ROHIT GUPTA

+61 432 805 123 | rohitgupta5533@gmail.com|
https://www.linkedin.com/in/rohit-gupta-88b763136

Data Scientist with 5 years of commercial work experience executing data driven solutions to increase efficiency, accuracy of internal data processing, analyzing, and interpreting large datasets, developing machine learning models, and performing data management tasks. My interpersonal skills have been developed by managing and interacting with clients and understanding their requirements. Proven experience in designing, implementing and harnessing the power of raw data to deliver insights and solutions to challenging business problems.

SKILL SUMMARY

- **Big Data Technologies:** Data Lake - Qubole
- **AWS:** S3, RDS, EC2, AWS Cli, Sagemaker
- **GCP:** Google Compute Engine, Vertex AI, Google Cloud Storage
- **LLM** : OpenAI-GPT4, Bing, Bard
- **Data Warehouse Tools:** GCP BigQuery
- **Data Science Technologies:** Machine Learning, Visualization libraries like Plotly & Seaborn, Processing libraries like Pandas, NumPy, Deep Learning Architectures, PCA, Statistical Analysis, A/B Testing, Causal Impact Analysis
- **Programming:** Python, R, Shell Scripting
- **Database:** SQL (MySQL), PostgreSQL
- **Tools:** Git, Visualization – PowerBI, Google Data Studio, Tableau, Databricks for machine learning deployment - MLFlow, Matillion, Supermetrics

WORK EXPERIENCE

AI & Data Science Tutor
Superprof, Remote **January 2020 - Present**

- Delivered online tutoring sessions in data science, creating personalized and engaging learning experiences.
- Produced high-quality educational content, including interactive exercises and tutorials, to facilitate understanding and retention.
- Provided mentorship to students considering careers in AI and Data Science, offering guidance on academic and professional pathways.
- Actively sought and responded to student feedback to improve teaching methods and materials, ensuring that each tutoring session was effective and enjoyable.

Data Scientist **December 2021 - Present**
In Marketing We Trust, Sydney, Australia (Remote)

- Currently running a project on Generative AI for an international travel website to generate content at scale using custom RAG architecture to incorporate client requirements with the help of frameworks like Langchain, vector databases, mysql database.
- Implemented multithreading in Generative AI script to speed up content generation process.
- Crafted and optimized AI prompts to produce content that aligns with client branding and audience engagement goals.
- Worked closely with clients to understand their requirements and integrate their feedback into the AI model's development for better-targeted content solutions.
- Content Production speed increased manifolds and cost of generation reduced significantly for the client with ~3% statistically significant upliftment of traffic due to fresh content.
- Developed a GPU based clustering model in Python to identify the control and test group for A/B testing on the client website traffic.
- Built a pipeline in R to perform data preprocessing of raw data and apply statistical analysis - causal impact on the clustering group to identify the significance in the performance of the changes to improve the efficiency and upliftment of traffic.
- Evaluation of the experiment results using Causal Impact Analysis
- Increased the performance of the traffic on the client's website by ~13% by optimizing internal linking.
- Developed an automation processing using ETL tool Matillion to extract the raw data from Google , Facebook and Instagram API's, transform the data as per requirement and load the data into BigQuery, which then was utilized for dashboards built in Tableau and Google Data Studio.

Data Scientist **April 2021 – November 2021**
Bondi Labs, Melbourne, Australia

- Research on appropriate hardware solutions to carry out the data collection process in meat processing plants.
- Performed data cleansing, statistical and EDA data analysis using Python.
- Addressed business problems through Statistical Learning methods and deep learning-based computer vision and Machine Learning techniques.
- Implemented UNET, MASK-RCNN models with Resnet architecture for image semantic segmentation model using TensorFlow

Junior Data Scientist **June 2016 – July 2019**
Ericsson Global India, Noida, India

- SQL for data extraction and management for all router information, starting from their establishment to their decommissioning.
- Develop data insights using Power BI and advanced excel.
- Build and improve reporting that helps the stakeholders understand the performance of their business segment and identify action points.
- Developed complex predictive models using advanced statistical and machine learning tools to forecast the decommissioning of the network trunks, servicing of the network routers.
- Analyze model performance, customer behavior and market conditions in order to provide key insights to business stakeholders.
- Analyze the client's business requirements and processes through document analysis, workshops, and workflow analysis.

EDUCATION

Masters of Data Science **July 2019 – June 2021**
RMIT University, Melbourne,
Australia GPA: 3.7/4
Key Courses: Machine Learning, Data Visualization, Practical Data Science with Python, Database Concepts, Statistics, Data Preprocessing, Big Data Processing, Machine Learning.

BTech in Electronics and Communication **August 2011 – May 2015**
CIEM, Kolkata, India
GPA: 8.53/10

ACADEMIC PROJECT

Classifying Nuclei Cell type and predict isCancerous or not
RMIT University – Computational Machine Learning

- The objective of the project is to use deep learning concepts to identify the type of cell using the images and also predicting whether the cell is isCancerous or not.
- Developed VGG architecture for Convolutional Neural Network model and Multi-Layer Perceptron model for classifying the images.
- Implemented the concept of oversampling using data augmentation techniques to increase the number of images for cell type with minimum images.
- Achieved a model accuracy of 0.74 on the test data with a loss close to 0.64 and f1-score of 0.69 using the hyperparameter tuning on the deep learning model.

COVID19 Monitoring, Risk Assessment and Resource Management
RMIT University – Case Studies in Data Science

- The objective defined is to develop COVID19 dashboard for Monitoring, Risk Assessment and Resource Management as a part of a University project
  https://bijothomas.shinyapps.io/Covid19_Dashboard/
- To provide a solution, a dashboard showing various visualizations projecting the impact of Covid on individual country, state and district level and a prediction which takes in several inputs and will then forecast the resources that are needed to manage patients
- Forecasting is done using the prophet forecasting library in R.
- Clustering is done to shortlist types of underlying disease that a patient might have and are more vulnerable to covid-19

Prediction of Heart Disease

- Main objective of the project was to clean the data, analyze and find insights, perform feature engineering, build machine learning models to predict if the incoming patient with vital readings is likely to have heart disease or not.
- Performed feature engineering and introduced 4 new features which increased the performance of the model by 3%.
- Use of hyperparameter tuning, on 4 different machine learning algorithms, to find the optimal parameters which gives the accurate result.

REFEREES

Will be provided upon request.